Developing Cloud-Based Tools for Big Neural Data

The University of Pennsylvania

PI: Joost B. Wagenaar Grant Number: 1-K01S25436-01

Big Data has the potential to dramatically advance the electrophysiology biodata sciences in similar ways that it has transformed Genetics. Differences between these two areas dictate separate approaches to apply Big Data tools, and methods in order to provide successful assets to the research community. For one, neural datasets are very heterogeneous by nature. The data is difficult to interpret without knowing specifics about the data acquisition protocol, the experimental paradigm and the physiological state of the recorded subject. Many neural datasets are complemented with complex meta-data sets, which should be an integral component in any effort to integrate and share these data with other researchers. The goal of this project is to develop novel, generalizable Big Data tools to facilitate cloud-base analysis of complex multi-scale neural data. Epilepsy research will be used as a specific use case to guide the development of the tools. A cohort of established senior investigators performing epilepsy research will use and validate these tools in their laboratories. Epilepsy research is currently limited by its narrow focus on single models (animal or human) in individual centers and laboratories. Just as Genetics was revolutionized through Big Data techniques, so too can Epilepsy research be transformed through novel approaches to standardize, share, and mine data across groups of investigators. Over the past several years I have co-developed a NINDS funded cloud-based data platform, ://ieeg.org, giving me a central role in developing Big Data solutions for neural data, such as customized data sharing, large-scale cloud-based data analysis, and search and interrogation techniques for complex data and metadata. My scientific objectives for this project are: (1) to develop generalizable tools to curate, analyze, and interrogate multi-scale neural data, and (2) to create a platform that will galvanize a research community focused on sharing data, and methods to advance Big Data research in the basic and translational neurosciences. Equally important to this proposal, I present a training plan to prepare me for an academic career focused on Big Data in the neurosciences. This plan supplements my background in bioengineering and statistical modeling of neural data with broader data-science expertise in data integration and machine learning, and deeper domain knowledge of the clinical neurosciences. I have assembled a group of collaborators, basic investigators and clinician scientists, who will use the tools developed in this project to analyze and validate their data and methods. I will use the results of this project as the foundation for a R01 Grant application, in which I will expand the developed platform and tools to target other research domains (TBI, Emergency Care, Cardiac), as well as integrate other data-modalities such as Imaging, and Genomics. OMB No. 0925-0001/0002 (Rev. 08/12 Approved Through 8/31/2015) Page Continuation Format Page PUBLIC HEALTH RELEVANCE: The goal of this proposal is to advance Big Data research in the neurosciences by developing tools and techniques to interrogate electrophysiology data sets from animal models of human neurological disorders. Development of these tools requires close collaboration between domain experts in Neuroscience, Machine Learning, Statistics and Computer Science. When developed, this platform and these tools will allow investigators to share, collaborate, annotate, standardize and analyze large, complex, multiscale data sets that are a crucial first step in advancing this field.